Manual de utilização do módulo Scanner

A SPHINX propõe uma solução de leitura automática de questionários por scanner. O Sphinx Ambiente Multimídia com opção Scanner é um sistema que oferece, além de todas as funcionalidades existentes no Sphinx Ambiente Multimídia, outras opções avançadas para apresentação do questionário e para leitura automatizada das respostas objetivas por scanner. Ele permite a impressão dos questionários em formato formulário, o posterior reconhecimento automático das respostas e a geração de arquivos com as respostas Sphinx®.

- **Simples:** O Sphinx Ambiente Multimídia com opção Scanner funciona como um novo modo de entrada de dados no Sphinx®. Um painel de comandos permite ao usuário realizar a leitura ou a consulta de questionários já lidos.
- **Rápido:** A velocidade de leitura dos questionários depende da rapidez do scanner e da capacidade do carregador; ela pode variar de 10 a 500 questionários por hora, isto se utilizado com os scanners recomendados (ver mais adiante).
- **Eficaz:** A taxa de reconhecimento automático é superior a 99%. Um ambiente amigável permite o controle dos resultados e a digitação das respostas às questões abertas.
- **Potente:** Reconhecimento de escalas, identificação de questionários nominativos, questionários frente-e-verso ou em várias páginas, reconhecimento de números.

Cenário de utilização do Sphinx Ambiente Multimídia Scanner:	
Criação do questionário	O questionário é redigido no Sphinx® com todas as facilidades do sistema (bibliotecas, desvios, partes, grupos, etc.).
Paginação e Impressão	Este questionário é aberto no Sphinx Ambiente Multimídia com opção Scanner e então um Formulário Scanner é criado segundo o estilo que você escolher ou definir. Todos os estilos podem ser personalizados. Diversas opções de impressão integradas ao módulo permitem uma formatação organizada (inserção de logotipo, disposição, fontes, colunas, etc.). Os formulários scanner serão em seguida impressos e reproduzidos (cópias) para serem então aplicados.
Leitura de questionários	Os questionários respondidos são colocados no carregador de documentos do scanner e lidos automaticamente. A leitura pode ser controlada e completada (áreas manuscritas) antes da validação do conteúdo.
Tabulação dos resultados	Um simples clique na tabulação automática apresenta imediatamente os resultados das tabelas simples. E ainda todas as facilidades de estatística e gestão de dados que o Sphinx® disponibiliza.

Importante: este documento não tem a pretensão de detalhar o uso do Sphinx Ambiente Multimídia, mas somente a parte referente ao uso do Scanner. As demais funções do Sphinx Ambiente Multimídia são apresentadas em documentações disponíveis no site da SPHINX Brasil.

Sumário

1. Apresentação geral

O Sphinx Ambiente Multimídia com opção Scanner é um sistema complementar ao Sphinx®, que oferece opções avançadas para apresentação do questionário e para leitura automatizada das respostas objetivas por scanner. Ele permite a impressão dos questionários em formato formulário, cujos aspectos de apresentação devem ser padronizados para o subsequente reconhecimento do questionário e de suas respostas pelo scanner. Depois de respondidos, os formulários preenchidos podem ser lidos por scanner, com identificação automática das respostas marcadas e geração de respostas em arquivo Sphinx. As respostas marcadas serão reconhecidas automaticamente, desde que os questionários que serão digitalizados tenham sido necessariamente elaborados e impressos a partir daquele mesmo exato arquivo (sem nenhuma alteração posterior).

Descrição do material e sistema necessários:

Os elementos necessários para a utilização da leitura automática de questionários com scanner são os seguintes:

- **O software Sphinx (de concepção do questionário):** o Sphinx será utilizado para elaborar o questionário/formulário, e deve estar instalado no computador do responsável pela pesquisa. O uso do scanner estará atrelado a instalação do Sphinx Léxica naquela máquina.
- **O sistema de leitura por scanner:** este sistema faz parte do Sphinx Ambiente Multimídia com opção Scanner, que deve estar instalado na máquina ligada ao scanner.
- **O scanner:** deve ser um scanner de produção, rápido (mínimo 10 páginas por minuto), equipado com um carregador fixo de documentos e compatível 100% com a norma Twain 32 bits.

Atenção! O Sphinx Ambiente Multimídia com opção Scanner não funciona com qualquer tipo de scanner. São necessários modelos específicos. Entre em contato conosco (suporte@sphinxbrasil.com.br) para conhecer os modelos atualizados.

2. Instalação do Sphinx Ambiente Multimídia com opção Scanner

A instalação do scanner deve ocorrer junto ao sistema Sphinx Ambiente Multimídia com opção Scanner. Desta forma, além de possibilitar o uso de todas as funções do Sphinx Ambiente Multimídia, a opção Scanner permitirá a impressão de formulários específicos para reconhecimento no scanner, além de funções para «pilotar» o scanner e realizar a leitura dos questionários preenchidos.

As principais diferenças entre o Sphinx Ambiente Multimídia e o Sphinx Ambiente Multimídia com opção Scanner são:

- No estágio de definição de formulários, uma opção «Scanner» diferenciada é disponibilizada, permitindo criar formulários específicos para reconhecimento via scanner.
- Estilos diferentes, específicos para o uso de formulários scanner, são disponibilizados.
- Os formulários scanner criados e impressos para preenchimento e posterior reconhecimento deverão ser os mesmos que serão utilizados no momento da leitura via scanner.
- A leitura e o reconhecimento serão feitos a partir do estágio «Digitar», onde uma função ativará a leitura óptica propriamente dita. O Sphinx Ambiente Multimídia com opção Scanner deve estar instalado no mesmo computador onde estará conectado o scanner.

3. Princípios gerais de funcionamento

No documento impresso, o entrevistado marca uma ou mais casas de cada questão fechada e escreve uma resposta (número ou texto) para as questões abertas. Questões abertas, se existentes, deverão ser digitadas manualmente durante ou após a digitalização dos questionários.

Os documentos são introduzidos no scanner. O sistema faz a aquisição das imagens da página. Esta imagem será em seguida examinada para reconhecimento das áreas de digitação, e posterior identificação do conteúdo.

Para encontrar automaticamente as áreas de digitação, o Sphinx Ambiente Multimídia com opção Scanner vai registrar sua posição exata durante a impressão. Infelizmente, a mecânica do scanner não permite obter uma imagem exatamente conforme a imagem original: de fato, por vezes, as folhas introduzidas no scanner podem ser aspiradas (inseridas no scanner) um pouco atravessadas, e a velocidade de passagem das folhas nem sempre é constante. É então necessário imprimir um quadro ao redor do formulário, o qual será feito anteriormente, antes da aplicação dos questionários, ou seja, na formatação dos formulários. Este quadro será facilmente detectado e o sistema poderá retificar a imagem. Ou seja, nada deve ser escrito fora do quadro, pois não será reconhecido.

Por outro lado, o sistema deverá ter uma informação sobre a página a analisar: trata-se de uma folha correspondente à pesquia? Qual é o número da página (se o questionário comportar várias páginas)? Estas informações, chamadas de códigos de página, aparecem sob a forma de casas brancas e pretas, na base do documento impresso (rodapé).

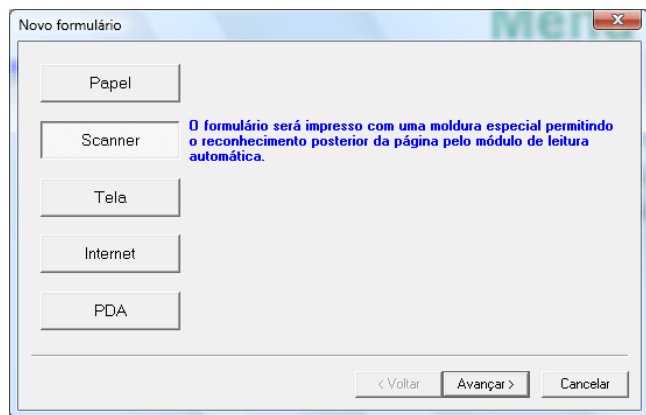
Quando a página do questionário for identificada, a detecção de respostas é iniciada. Cada área de resposta será analisada. Para as casas de marcar, olha-se no interior de cada casa: ela está marcada? ela está rasurada? Para tal, mede-se o percentual de preto no interior da casa até os limites da casa: abaixo de um determinado percentual, considera-se que se trata de um risco qualquer, se for superior a outro limite especificado, considera-se que a casa está rasurada, se estiver entre os limites inferior e superior, considera-se como uma resposta (a casa estará marcada). Quando todas as questões tiverem sido lidas, o sistema mostra na tela as respostas dentro da mesma paginação que o documento tiver sido anteriormente impresso. Para as questões fechadas, as categorias marcadas aparecem como uma cruz. Para as questões abertas, é a imagem digitalizada da área de resposta que será exibida. O usuário pode então, de forma manual, digitar as respostas das questões abertas (que não serão automaticamente reconhecidas).

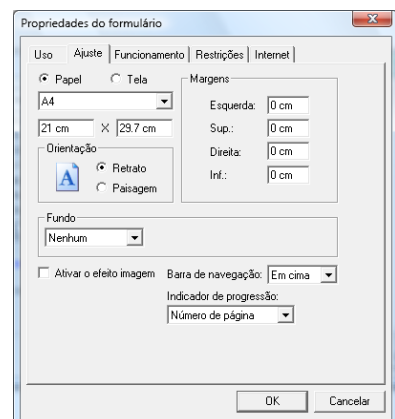
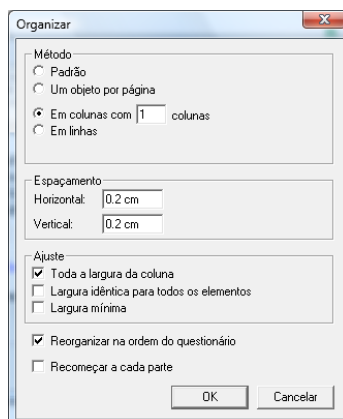
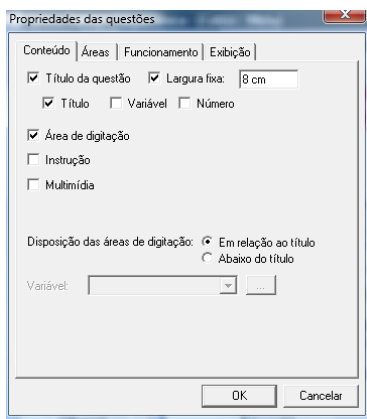
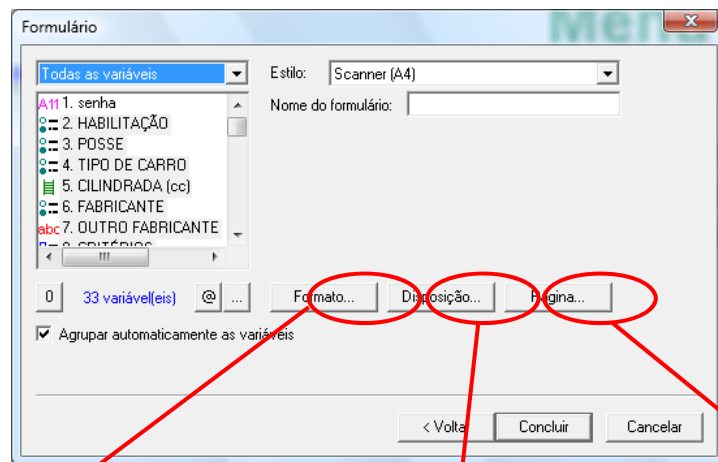
4. Preparação dos documentos

4.1 Criação dos formulários

Os formulários Scanner são criados a partir do Sphinx Ambiente Multimídia com opção Scanner.

Para tal, abre-se no Sphinx Multimídia com opção Scanner um questionário anteriormente definido no Sphinx® (arquivo .QUE), e em seguida, estando no estágio de Definição de Formulários, acessa-se o menu Formulários, opção Novo Formulário.

Na janela que se abre, clicar no botão Scanner, e em seguida clicar no botão Avançar.

Em seguida, será apresentada uma tela com várias opções de definição do formulário como a escolha de variáveis, estilo, formato, disposição dos elementos no formulário e formatação da página, etc.

Cada uma das opções abre um conjunto de itens que permitirão definir a apresentação do formulário que está sendo criado.

A partir do botão **FORMATO** é aberta uma janela de definição das propriedades das questões, onde é possível configurar o conteúdo que será apresentado (título, largura, variável, número, instruções, etc.), as áreas de resposta (numeração das categorias, formato das casas de marcar, número de colunas dos itens de resposta, tamanho, etc.), o funcionamento e o desenho (fontes, tamanhos, alimento, cor, bordas, etc.).

A partir do botão **DISPOSIÇÃO** é possível organizar a disposição dos elementos no formulário, número de colunas em que as questões são apresentadas, espaçamento entre colunas, ordem, partes, larguras, etc.

A partir do botão **PÁGINA** são apresentadas as opções do scanner (margens especiais, código das páginas, numeração especial das páginas, etc.), as opções de funcionamento (tipo de marcação, controle de digitação, controle de realização, etc.) e o tamanho (formatação de fontes, cores, margens da página, etc.)

Outras opções para configurações especiais serão apresentadas na sequência deste documento, (no item Precauções Indispensáveis), como por exemplo, as operações que devem e as que não devem ser feitas para otimizar a tarefa de uso do scanner e assegurar confiabilidade no procedimento.

Importante! Se os questionários forem preenchidos pelos próprios respondentes, é necessário que eles sejam advertidos da leitura automática por scanner, incitando-os a marcarem suas respostas de forma visível. Pode-se, para tal, apresentar uma imagem que mostra os tipos de marcas corretas e os tipos de marcas incorretas num formulário. Uma imagem (ConsigneScan.gif) contendo exemplos de marcação de casas pode ser encontrada no diretório onde o Sphinx estiver instalado, subdiretório SphinxV5\Multimédia\Formes

4.2 Impressão e reprodução dos formulários

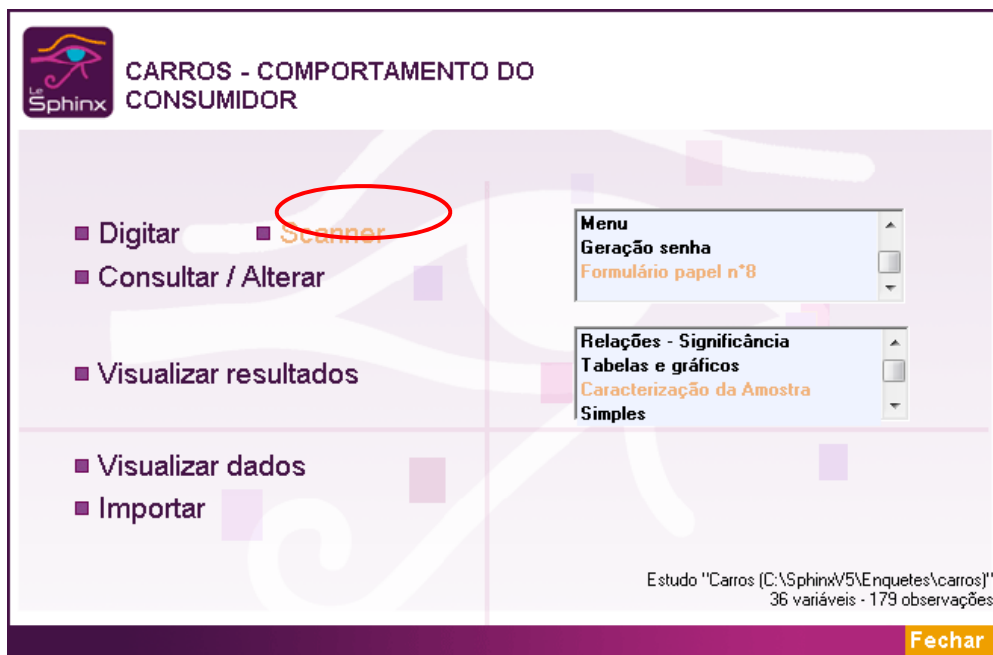
Tendo criado o seu formulário no Sphinx com opção Scanner, você poderá aplicá-lo sob 2 formas:

- **Com fotocópia tradicional:** o formulário será impresso numa impressora de boa qualidade (em branco e preto ou colorido), e em seguida tal impressão poderá ser reproduzida em xerox, para então ser aplicada.
  Atenção! É importante tomar cuidado com a qualidade da fotocópia porque, por exemplo, rasuras ou manchas no formulário poderão confundir o sistema no momento do reconhecimento, impedindo a leitura correta. Recomendamos realizar testes com um formulário reproduzido antes de realizar a reprodução em grande quantidade.

- **Com cópia digital:** o formulário poderá ser transmitido sob a forma de um arquivo digital. O formato normalmente utilizado é o PDF do Adobe. Para gerar um arquivo no formato PDF a partir do Sphinx Ambiente Multimídia Scanner, é necessário possuir um programa que permita a criação de tais arquivos. O programa Acrobat da Adobe é recomendado. Se ele estiver instalado no mesmo computador em que você estiver trabalhando com o Sphinx, a geração das cópias digitais será simples: bastará imprimir o formulário normalmente a partir do Sphinx Ambiente Multimídia com opção Scanner, escolhendo para tal a impressora virtual Adobe PDF Writer disposta na lista das impressoras disponíveis. Em seguida, será criado um arquivo reproduzindo fielmente o formulário. O respondente poderá então imprimi-lo de fato em papel e responder ao formulário, enviando em seguida ao pesquisador.

5. A leitura automática

5.1 O ambiente

Para iniciar a leitura dos formulários por scanner, execute o Sphinx Operador com opção Scanner. Na tela principal, escolha a opção «Scanner» que somente estará presente se o formulário tiver sido definido como um formulário do tipo scanner, e não como dos outros tipos de formulários disponíveis (Internet, Papel e Tela Multimídia).

Em seguida, o ambiente de leitura será apresentado na tela, com uma mensagem indicando que é momento de inserir os questionários na bandeja do carregador de documentos, clicando, em seguida, no botão «Scanner».

5.2 Lançar a leitura dos formulários

- Clique no botão «Scanner» para ler sucessivamente todas as folhas colocadas na bandeja do carregador de documentos do seu scanner;
- Desejando ler somente uma folha (por exemplo numa fase de testes), clique ao mesmo tempo a tecla CTRL e o botão Scanner;
- As folhas serão lidas e a lista das páginas lidas será incrementada aos poucos. Cada linha da lista das páginas informará o número da observação e a página.

Após a leitura, a primeira página lida será exibida na tela. Não se trata de uma imagem da folha digitalizada, mas sim o questionário reconstituído após reconhecimento das casas marcadas. Somente as áreas de questões abertas corresponderão a imagem original.

Se incoerências forem detectadas, uma linha vermelha indicará tal área. Essas incoerências são os controles de digitação de respostas não respeitadas no formulário em questão. Por exemplo, se uma questão declarada como obrigatória no questionário não tiver sido marcada. Ou então se várias categorias tiverem sido marcadas para uma questão definida como fechada única. O software tentará realizar uma correção inteligente e serão as casas mais prováveis que serão consideradas.

É então possível mostrar a imagem completa da página para tirar a dúvida numa questão ambígua. Por exemplo, pode-se perceber que duas casas tenham sido marcadas, mas que o respondente indicou com uma flecha aquela que deveria ser considerada (a boa resposta), algo que o computador seria incapaz de entender.

Se vários erros forem detectados na página, será impossível mostrar os «Diagnósticos de leitura» para ter a lista detalhada. Quando uma página com erros for visualizada, ela aparecerá na cor Magenta na lista das observações. Desta forma, será possível saber quantas ambiguidades deverão ainda ser corrigidas. Ao possuir uma lista sem observações na cor Magenta, ela estará pronta para ser efetivada como respostas à pesquisa.

5.3 Corrigir e completar

Se ambiguidades forem detectadas, uma boa técnica é alternar entre a imagem completa da página e a exibição das respostas reconhecidas (escolhendo a opção na lista drop-down ou utilizando o atalho F6). Se a leitura tiver que ser corrigida, basta clicar com o mouse sobre a categoria correta para modificar a página.

Para as áreas de digitação de questões abertas, visualiza-se a porção correspondente da imagem. Basta clicar para fazer aparecer uma caixa de texto dentro da qual será fácil digitar com o teclado as respostas indicadas. Pode ser mesmo interessante, já nesse estágio de entrada das respostas, iniciar uma reformulação para facilitar a análise posterior.

Se a opção de reconhecimento automático de escrita estiver instalada, será possível alternar (tecla F8) entre as Respostas reconhecidas ou digitadas e a Imagem das áreas de digitação para corrigir possíveis casos de erros de leitura.

5.4 Erros de leitura

Se for impossível a leitura do formulário, uma mensagem será exibida informando a razão do erro e a leitura será interrompida. Por exemplo, a mensagem pode ser: «Impossível ler o quadro», ou então «Impossível ler os códigos de página», ou «Número de pesquisa incompatível». Uma mensagem vai lhe propor visualizar uma superposição da imagem produzida pelo scanner e os elementos esperados e detectados. Será possível, talvez, perceber algum erro no carregamento das folhas, como por exemplo, se a página introduzida estiver no lado errado ou que uma sobrecarga esteja impedindo a leitura correta, ou mesmo ainda que o formulário colocado no scanner não corresponda à pesquisa em questão.

Mais para frente neste documento, no item Precauções Indispensáveis, você encontrará a lista dos obstáculos que podem impedir o sistema de ler corretamente a página, e as maneiras de como evitá-los ou contorná-los.

Quando ocorrerem erros, a melhor coisa a fazer é deixar de lado a página com erros e relançar a leitura com as demais páginas do lote de formulários correspondente à pesquisa. As páginas não lidas serão repassadas no final após verificação ou eventual retificação. Se a leitura se mostrar definitivamente impossível, o questionário (observação) será digitado manualmente.

Pode-se assim chegar a erros que tenham sido detectados na leitura de uma questão que ao ser observada ao visualizar-se a imagem completa da página, aparentava ser normal. Escolhendo a opção Áreas de leitura na lista drop-down, é possível mostrar a imagem contendo, para cada casa de marcar, a casa correspondente esperada (em vermelho) e a casa detectada (em azul), combinando o percentual da superfície da casa que foi escurecida. É possível também alterar os limites máximos e mínimos admitidos para esse percentual.

5.5 Aplicar um valor a todas as folhas do pacote

É possível aplicar um mesmo valor para uma dada questão a todas as folhas de um lote. Exemplo:

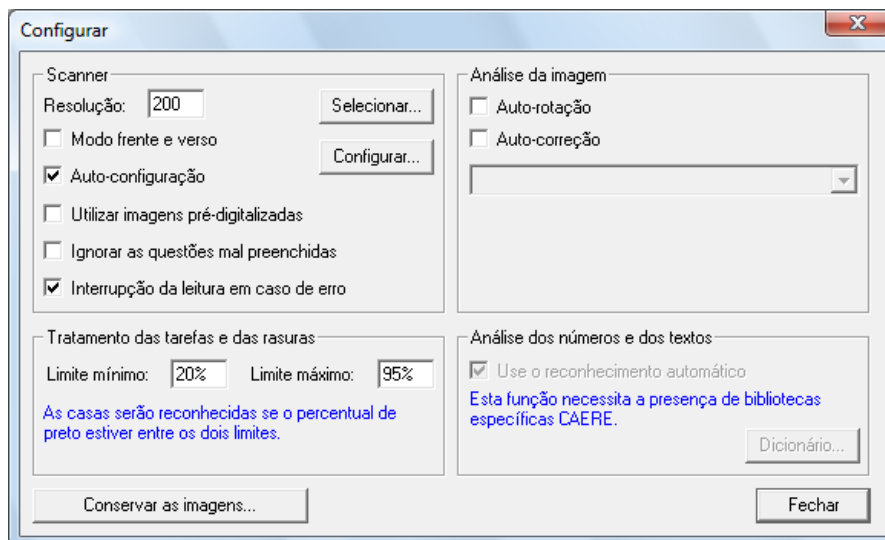
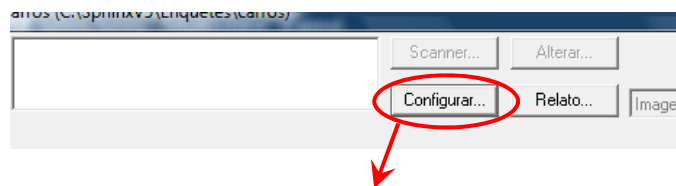
- Os questionários são do tipo questionário de satisfação de um estabelecimento de saúde com diversos tipos de serviços.
- estudo deve permitir comparar os serviços entre eles.
- serviço deve ser uma das variáveis definidas na lista de questões.
- Esta variável não pode ser uma questão fechada porque a quantidade de categorias (neste caso, os diferentes tipos serviços da clínica de saúde) é muito grande.
- No questionário, a variável serviço é associada a uma questão do tipo aberta código.
- Esta informação entra sob a forma manuscrita e o scanner não poderá reconhecê-las automaticamente.
- Se os lotes de questionários que serão inseridos na bandeja do scanner estiverem classificados por serviço, não é necessário digitar o código do serviço para cada questionário.
- Após cada leitura de um lote de questionários, para afetar um mesmo valor a uma questão, clicar no botão **Modificar** e em seguida no botão **Afetar**. Escolher a variável desejada (neste exemplo, a variável serviço) e indicar o valor que será atribuído a todas as folhas do lote (por exemplo, valor número 4, correspondente ao serviço «berçário» realizado na clínica de saúde).

5.6 Aceitar ou recusar um lote de folhas

Toda a entrada de dados por scanner é realizada a partir de lotes de questionários. Entende-se por lote de folhas cada pacote de questionários inserido na bandeja do scanner, a cada vez. Depois de finalizada a leitura das folhas de um lote de questionários, ele deve ser gravado ou anulado:

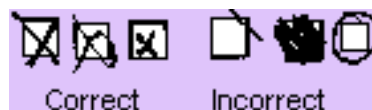
- **Gravação de um lote:** as observações (questionários) serão definitivamente adicionadas ao arquivo de respostas da pesquisa.
- **Anulação de um lote:** é possível anular um lote de folhas. Num caso em que o carregador lançar mais de uma folha ao mesmo tempo, por exemplo, é difícil identificar quais são, de fato, as folhas que foram efetivamente lidas. Em casos desse tipo, é preferível recusar o lote e reintroduzir as folhas no carregador e reparar folha a folha.

6. Regulagem do scanner

A regulagem do scanner é realizada a partir do botão « Configurar » disposto na tela principal de leitura de questionários por scanner.

6.1 Regulagem dos limites de reconhecimento

Os limites mínimo e máximo são expressos em percentuais. A partir do percentual mínimo, o sistema considera a casa como marcada. A partir do percentual máximo, o sistema considera a casa como rasurada (anulada).

O limite de 20% permite absorver as pequenas marcas provocadas pela impressora ou mesmo durante o preenchimento do questionário pelas pessoas. O limite de 95% permite tratar as rasuras, ou seja, se um respondente desejar anular uma marca já feita para marcar outra, basta preenchê-la na sua totalidade e o sistema reconhecerá como sendo anulada. É importante precisar aos respondentes estas possibilidades de marcação.

Para bem regular o seu scanner e o reconhecimento das marcações, recomendamos realizar alguns ensaios para bem definir os limites desejados. Após a leitura de uma página, a opção Áreas de leitura permite ver os percentuais reparados para cada casa marcada. Valores de 15% a 25% para o limite mínimo e de 85% e 100% para o limite máximo são considerados normais.

6.2 Regulagem do método de reconhecimento (resolução, frente e verso, auto-rotação, auto-configuração)

Resolução: 200 (pontos por pixel) é uma boa regulagem entre a rapidez de aquisição de imagem e sua resolução. Em certos casos, pode ser necessário aumentar essa resolução para 300 (por exemplo, se os riscos das marcações das casas forem muito finos e pouco fortes).

Frente e verso: certos scanners permitem digitalizar as páginas frente e verso de uma folha de uma só vez. A vantagem não é somente ganhar velocidade, mas igualmente facilitar enormemente o tratamento das folhas: sem riscos de misturar páginas, sem necessidade de ficar organizando os questionários em ordem das páginas, etc. Entre em contato conosco para conhecer a lista atualizada de scanners compatíveis.

Auto-rotação: o software supõe que a folha tenha sido introduzida no sentido natural, ou seja, o alto da página primeiro. Se os elementos necessários não forem detectados (ver em explicações sobre os códigos de página, apresentados mais para frente neste documento), um erro será assinalado.

Marcando o item «Auto-rotação», indica-se ao software que ele deve tentar reorientar automaticamente a imagem. Esta função é muito útil nos seguintes casos:

- Os formulários não foram todos dispostos no mesmo sentido e deseja-se automatizar a tarefa de colocá-los todos no mesmo sentido;
- Deseja-se poder, deliberadamente, inserir as folhas ao inverso, para facilitar sua introdução: por exemplo, se as folhas forem seguradas (trancadas), a deformação das páginas pode entupir o carregador que não conseguirá separar as diferentes folhas de um mesmo questionário;

- Os formulários foram realizados no formato paisagem e não existe um sentido evidente para a disposição dos questionários no carregador. Deixa-se então que o software determine o bom sentido.

Auto-configuração: quando todas as casas estiverem marcadas, o software posicionará automaticamente as regulagens do scanner: tamanho da folha de papel, seleção do carregador, etc. Pode-se desativar esta auto-regulagem, clicando no botão «Auto-configuração». Isto pode ser necessário quando, por exemplo, deseja-se realizar testes com um scanner não munido de um carregador de documentos.

6.3 Selecionar e configurar o scanner

Se vários scanners (ou drivers de scanners) forem instalados, o botão «Alterar» permite escolher o scanner que será utilizado para o reconhecimento dos questionários.

O botão configurar o «Scanner» leva ao diálogo de configuração do scanner fornecido pelo fabricante do scanner ou de seu driver. Esse é ambiente no qual será possível realizar as regulagens necessárias para aumentar ou diminuir o contraste e a luminosidade, de forma a obter uma imagem mais visível possível.

Se o questionário tiver sido impresso em cores, será necessário programar o tratamento das cores de base para somente reter os elementos úteis (ignorando, por exemplo, o fundo da casa se esta tiver sido impressa em cores).

Certos scanners solicitam regulagens particulares para obter a imagem adequada. Por exemplo, alguns scanners propõem a imagem em negativo se a opção Inverter não estiver marcada.

6.4 Corrigir as deformações devidas ao scanner

Os scanners normalmente utilizam um motor para carregar as folhas que estão dispostas na bandeja. Em certos scanners, o carregamento não possui uma velocidade constante e é necessário aplicar uma correção no momento da detecção das casas marcadas. Um jogo de correções pré-definidas é proposto para os scanners que recomendamos: escolha a linha correspondente na lista drop-down.

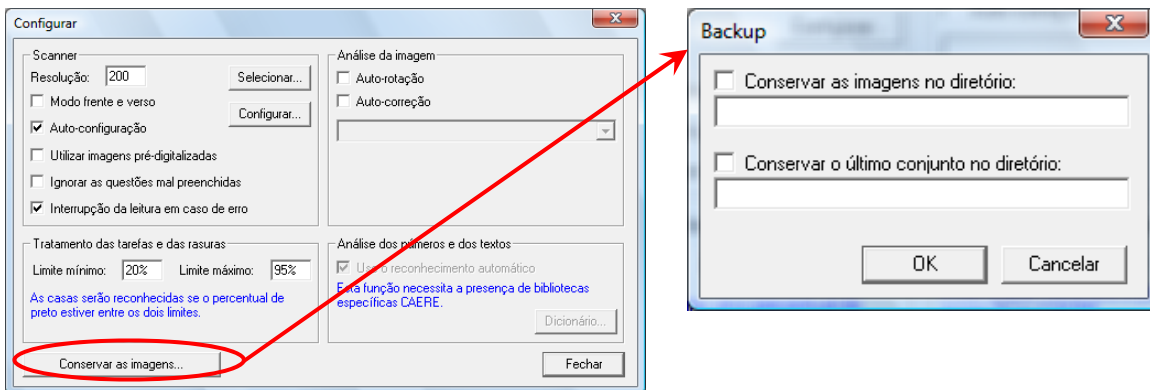
É possível regular seu scanner se ele não aparecer na lista ou mesmo ajustar um scanner se um uso intensivo ou um problema qualquer tiver modificado seu comportamento ou mesmo ainda se o documento tiver sido impresso numa folha especial (como um papel muito grosso, por exemplo). Clicar no botão «Nova correção» e indicar um nome a esse novo elemento. Clicar no botão «Modificar» para imprimir, digitalizar e depois corrigir a página de calibragem ou teste.

Para verificar se o scanner está corretamente ajustado, basta examinar as Áreas de leitura (opção da lista drop-down do ambiente de leitura). As casas esperadas devem estar quase exatamente superpostas às imagens das casas reconhecidas pelo scanner durante a leitura.

7. Funções avançadas

7.1 Gravação das imagens

As imagens são armazenadas temporariamente num diretório indicado no diálogo «Backup»,

acessível a partir do botão «Conservar as imagens». Se nenhum diretório for especificado, é o diretório «Imagens» da pesquisa em andamento que será utilizado e eventualmente criado.

Essas imagens serão conservadas para permitir a visualização da imagem digitalizada para fins de verificação e para permitir a digitação de questões abertas visualizando a área relacionada à imagem. Numa ótica «sem papel», é possível registrar as imagens. Atenção ao tamanho requerido para a gravação de uma imagem. Uma página no formato A4 ocupa em torno de 500 Kb.

7.2 Gravação do último pacote lido

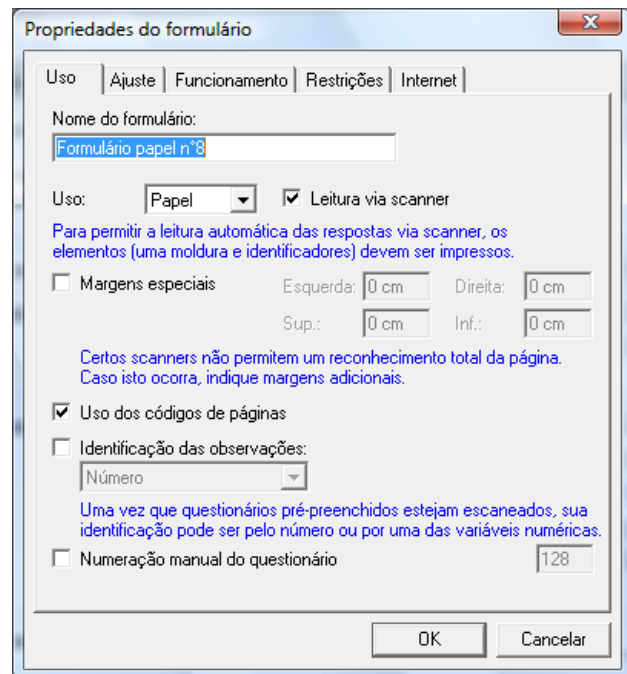
Uma opção permite gravar o último pacote (lote) de questionários digitalizado, não somente durante a pesquisa em andamento, mas também em outra pesquisa cujo nome for indicado. O objetivo desta opção é permitir obter uma visualização imediata do último pacote de questionários lido, sem a necessidade de recorrer aos originais.

7.3 Códigos de página

Os códigos de páginas são identificadores de formulários, apresentados sob a forma de quadradinhos pretos e brancos, localizados no rodapé do formulário.

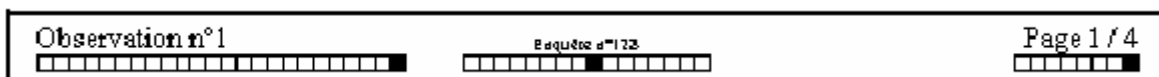
As configurações destes códigos são acessíveis durante o estágio de definição do formulário scanner.

Para tal, acessar o menu Formulários, opção Propriedades do Formulário, e clicar na ficha «Uso».

Na janela que será apresentada, estarão dispostos os itens capazes de definir ou personalizar os códigos de páginas de seu formulário, como o uso ou não dos código de página, o identificador da pesquisa (numeração manual do questionário), margens especiais, identificação especial das observações via uma questão específica, etc.

Existem 3 diferentes tipos de códigos de página, que são:

Identificador da pesquisa: quando o formulário for impresso, o software lhe atribuirá automaticamente um número (o qual é incrementado a cada nova impressão quando a paginação for alterada). Este número será registrado no formulário e inscrito no documento impresso sob a forma de códigos de página, localizado no centro do rodapé. O sistema verificará assim qual folha foi introduzida no scanner, e poderá saber se ela corresponde efetivamente ao formulário ativo da pesquisa aberta.

1 **Identificador da página:** quando o questionário possuir mais de uma página, um outro conjunto de códigos de páginas será apresentado, localizado à direita do rodapé, permitindo ao software diferenciar as diferentes páginas do formulário.

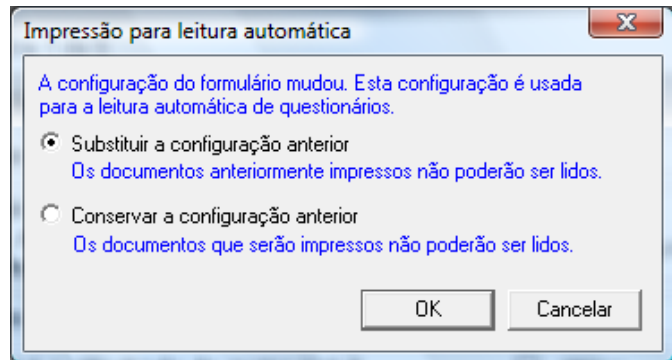
2 **Identificador do questionário (observação):** quando os formulários estiverem já pré-preenchidos com algumas informações referentes a cada respondente (em casos em que os respondentes já são conhecidos), a identificação da observação é representada sob um terceiro conjunto de quadradinhos, localizado à esquerda do rodapé.

Desejando, é possível imprimir formulários sem a identificação de códigos de página: basta desmarcar o item «Uso de códigos de página». Tal opção pode ser utilizada em casos nos quais o formulário não comporta mais do que uma única página, e que o usuário assegura por ele mesmo a conformidade da página lida e do formulário esperado. A estética do formulário será alterada, e poderá apresentar um risco de erro. Por exemplo, se a folha for introduzida ao inverso, ou seja, com a base da folha em primeiro lugar, o quadro da

página será bem detectado, mas os códigos de página não serão reconhecidos, gerando um erro de reconhecimento. Se os códigos de página estiverem ausentes, o sistema não poderá identificar o erro de introdução da página, as detecções das marcas sairão erradas.

Vimos que a **identificação da pesquisa** é incrementada automaticamente a cada impressão diferente do formulário.

No momento da impressão, o software verificará que a configuração do formulário mudou e solicitará a confirmação da validade da nova formatação para uma mesma pesquisa.

Ao substituir a configuração anterior, os antigos formulários já impressos não poderão mais ser lidos.

Ao conservar a configuração anterior, os novos formulários que forem impressos não poderão ser lidos. Isso porque a identificação da pesquisa não será mais a mesma.

Para estes casos em que se deseja reimprimir a mesma pesquisa com uma formatação diferente, mas deseja-se conservar a mesma identificação, é possível personalizar o número de identificação da pesquisa durante a definição do formulário, clicando no item «Numeração manual do questionário» e definindo o número desejado.

7.4 Margens especiais

Existem certos scanners que "comem" a borda da folha e do quadro identificador da pesquisa, isto em função das mesmas estarem dispostas muito perto dos limites da folha. Nestes casos, corre-se o risco da página não ser reconhecida pelo sistema.

A solução é forçar uma redução da área de impressão, alterando os valores das margens inferior, superior, esquerda e direita do formulário. Para tal, basta clicar no item «Margens especiais» e definir os valores desejados.

Este problema não existe com os scanners recomendados (entre em contato conosco para conhecer os modelos atualmente recomendados).

7.5 Digitalizar no formato paisagem

Se os formulários forem impressos no formato paisagem, será necessário marcar o item «Autorotação» para indicar ao software que os códigos de página não se encontram na borda inferior (rodapé) da imagem digitalizada, mas sim numa das bordas laterais. O item «Auto-rotação» é disponibilizado a partir do botão «Scanner» disposto na tela principal de leitura por scanner (ver item 7.2).

7.6 Retificação da posição das casas

Se erros de leitura forem identificados, pode ser porque existem deformações da imagem. Nestes casos, é possível retificar manualmente a posição esperada das casas. Durante a leitura por scanner, na janela de definição das áreas de leitura, escolhe-se a opção «Modificar a posição esperada das casas» na lista drop-down. Para cada casa, será utilizado o mouse para deslocar a casa esperada (em azul) até uma completa superposição com a imagem da casa. Esta retificação será conservada para todas as próximas leituras de outros questionários.

8. Tratamento dos questionários identificados

Muitas vezes deseja-se aplicar uma mesma pesquisa para respondentes que já responderam anteriormente a mesma pesquisa. Pode ser o caso, por exemplo, de avaliação de uma empresa em 2 anos diferentes, levantamento sócio demográfico, ou confrontação de respostas a um mesmo produto em diferentes momentos. Um protocolo de questionamento frequentemente utilizado consiste em enviar questionários identificados para poder, em seguida, confrontar as informações recebidas com outras informações existentes numa base de dados ou num painel.

8.1 Identificação anterior das observações

Quando da reaplicação de um questionário ou painel, os indivíduos a serem interrogados são importados da base de dados já existente para uma nova pesquisa. As questões de identificação são então apresentadas para cada um dos indivíduos. Pode-se, então, no formulário, mostrar certas informações conhecidas, como por exemplo o nome da pessoa ou o número de sua identificação. A segunda etapa consiste em imprimir o conjunto dos formulários. No ambiente do estágio «Consultar/Alterar», pode-se imprimir o conjunto dos formulários identificados (com algumas questões já preenchidas, e outras que serão aplicadas).

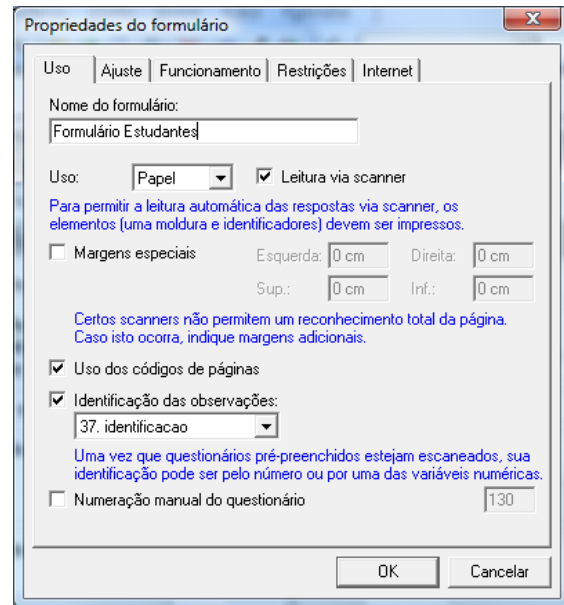
Nesses formulários, os códigos de página permitirão, na sequência, a identificação da observação.

Esta identificação pode ser feita pelo número de identificação do indivíduo. Não será possível modificar a ordem das gravações ou suprimir uma gravação sem que todas as observações sejam deslocadas.

Uma solução é ter no questionário uma variável numérica que permita identificar cada observação.

Marcando a casa «Identificação especial das observações» e escolhendo esta variável na lista, os códigos de página representarão o valor desta variável (cabe ao pesquisador assegurar a unicidade dessa variável).

No exemplo ao lado, da pesquisa Estudantes, utilizou-se a variável «Identificação» como chave de identificação do indivíduo.

Quando os formulários forem lidos, o sistema detectará automaticamente a observação (respondente) correspondente que irá documentar as novas observações lidas no formulário.

8.2 Identificação posterior das observações

Os questionários são reproduzidos no ambiente de impressão que dispõe de um sistema para imprimir uma informação diferente em cada exemplar, sob a forma de um selo. O encarregado dos estudos comunica ao ambiente a lista das quantidades/números que devem ser impressas em cada questionário.

O Ambiente Multimídia com opção Scanner permite tratar tais questionários graças à leitura automática das quantidades impressas. Uma variável correspondente ao número do dossier será criada no questionário e posicionada no formulário. É dentro da área de digitação correspondente a essa questão que o gerenciador de impressão deverá fazer aparecer os números. É importante cuidar que o gerenciador utilize uma fonte de caracteres fácil de ser reconhecida (uma fonte sem muitos detalhes, clássica, tamanho razoável, etc.).

Quando os questionários retornarem, eles serão lidos automaticamente (número do dossier e variáveis da pesquisa propriamente dita). Essa pesquisa, em seguida, será agrupada (fusão) com a base de dados contendo as informações correspondentes a cada número de dossier. O encarregado dos estudos poderá assim confrontar as respostas às questões da nova pesquisa com as informações já conhecidas para cada indivíduo (respondente, painelista, etc.).

9. Leitura de um formulário impresso sem o Sphinx com opção Scanner

Toda a tecnologia de leitura do Sphinx Ambiente Multimídia com opção Scanner é fundada na integração de 3 diferentes estágios do software: a formatação, a impressão e a leitura. As coordenadas das áreas de leitura são, assim, conhecidas em todos os estágios.

Entretanto, é possível, sob certas condições, ler automaticamente um formulário que não tenha sido impresso pelo Sphinx Ambiente Multimídia com opção Scanner. A condição essencial é a existência de um quadro contornando completamente o formulário. Se o quadro não existir, é possível mesmo assim efetuar

um desenvolvimento específico para simular o uso de outras partes geométricas estáveis do formulário. Isso somente funcionará se o volume de documentos pré-existentes for considerável.

Se o quadro existe, reconstitua no Sphinx Ambiente Multimídia com opção Scanner um formulário o mais parecido possível do documento já impresso que você possui (não esqueça de desmarcar a opção «Utilizar os códigos de página»). Em seguida, execute a leitura por scanner. O software provavelmente irá detectar erros de leitura, mas, utilizando a possibilidade de retificar manualmente o tamanho e a posição das casas esperadas, será possível tornar tais documentos legíveis.

10. Precauções indispensáveis

Será apresentado a seguir um conjunto de regras a observar para uma utilização satisfatória do modo de leitura por scanner. Essas regras dizem respeito a todos os aspectos da leitura por scanner, da elaboração do formulário à passagem dos documentos no scanner.

Definição: ZLA é uma Zona Lida Automaticamente, e pode ser um código de página, uma casa de marcar, uma área para reconhecimento óptico (escrita manuscrita ou impressão por uma máquina).

10.1 Definição do formulário

- É importante cuidar para conservar um espaço livre de no mínimo 1 mm ao redor das ZLA (atenção a sobreposição das questões).
- Não colocar cores de fundo para as ZLA.
- Utilizar a cor preta para o traço das ZLA.
- Não colocar sombreamento nas casas de marcar.
- Quando a elaboração do formulário estiver bem avançada, é prudente imprimi-lo e verificar que o quadro forma uma linha fechada e contínua ao redor do conjunto das questões na página, e que não existem ZLA parcialmente ou totalmente escondidas por outra questão.
- Quando o documento impresso estiver satisfatório, é importante preencher e digitalizar alguns exemplares, de forma a verificar que não existem problemas.
- Antes de sua difusão (aplicação), o formulário deve ser reproduzido. O formulário pode ser transmitido ao serviço de reprodução de duas formas: [1] um documento papel original, que é o resultado de uma impressão numa impressora de boa qualidade (preferencialmente laser). É uma solução válida para formulários preto e brancos; [2] um arquivo no formato Adobe Acrobat (PDF) acompanhado de uma impressão papel. Neste caso, é necessário possuir o software Adobe Acrobat Writer. Esta é a solução mais interessante porque se trata de um arquivo que já foi enviado à impressora (no formato PDF), e então estará livre de novas etapas de numeração do documento impresso. Além disso, permite também a separação das cores para uma reprodução em cores.

10.2 Impressão do formulário

Recomendamos sempre utilizar o mesmo computador e a mesma impressora durante a fase de impressão e testes dos formulários que serão utilizados. De fato, pode haver, por vezes, diferenças ou detalhes mínimos entre os sistemas de impressão ou entre as fontes utilizadas pelas impressoras. Tais diferenças podem gerar

alterações nas páginas e nas definições das casas, e, por isso, é importante tomar o cuidado e seguir o padrão de uma única impressora, de um único computador.

10.3 Reprodução do formulário

A qualidade da reprodução do formulário é muito importante. Aconselha-se utilizar um material do tipo profissional tomando cuidado com os seguintes aspectos:

- A margem ao redor do quadro da página deve ser em torno de 5 mm em torno dele.
- documento deve ser o mais fiel possível do original (sem distorções de imagem).
- documento deve ser isento de anotações, marcas ou rasuras, principalmente ao redor do quadro da página e nas ZLA.

Se, entretanto, anotações manuais tiverem de ser adicionadas por necessidades próprias da pessoa que aplica as pesquisas, estas deverão estar no interior do quadro da página e fora das ZLA.
Atenção às dobras do formulário: se o documento é destinado a ser dobrado (por exemplo, para ser colocado dentro de um envelope), é preferível realizar um teste previamente, verificando que as dobras não passam pelas casas de marcar.

10.4 Preenchimento pelos entrevistadores ou pelos entrevistados

Dois casos podem ser apresentados: ou a pessoa que preenche o formulário é um entrevistador a quem se pode ensinar algumas regras de base para a digitação das respostas, ou é o próprio entrevistado. O segundo caso deixa a porta aberta a todos os tipos de imaginação e pode eventualmente perturbar o bom reconhecimento das áreas de resposta.

Portanto, é importante aplicar as regras seguintes:

- Utilizar caneta preta, de traço nem muito fino, nem muito grosso.
- Para as casas de marcar, fazer uma cruz legível, tentando centralizá-las, e sem passar muito o traço para fora das casas.
- Se uma casa for marcada por engano, basta pintar toda ela para anulá-la.
- Para as áreas de reconhecimento óptico, respeitar um espaço livre de 1 mm no interior do retângulo de resposta.

10.5 Leitura automática pelo scanner

- Recomendamos preparar os documentos antes de introduzi-los no scanner:
  - Retirar os grampos.
  - Se o formulário tiver várias páginas, reagrupar as páginas de uma observação. As páginas não precisam estar ordenadas, mas devem estar agrupadas.
  - Se as páginas tiverem sido dobradas ou enroladas durante a aplicação da pesquisa, é importante colocá-las numa prensa de forma que elas retornem ao seu estado natural.
  - Folhear o maço das páginas para evitar que elas colem umas nas outras.

- Em seguida, colocar as páginas no carregador do scanner. A posição dos documentos no carregador, por padrão, deverá ser a posição normal de introdução (por exemplo, no scanner Fujistu M309XDC, coloca-se o cabeçalho do documento na parte de baixo do carregador). Entretanto, desejando facilitar a passagem dos documentos que estão num estado ruim ou deformados pela presença anterior de grampos, é possível introduzi-los no sentido oposto (cabeçalho em primeiro ⇔ cabeçalho por último). O sistema de encarregará de reorientar a imagem do documento. Nesse caso, o tempo de tratamento de uma página aumenta em alguns segundos. Atenção! Esta funcionalidade adicional aplica-se somente para formulários que possuem códigos de página.
- A quantidade de documentos que podem ser tratados a cada lote depende da capacidade do carregador de documentos do scanner e do estado das páginas a tratar.
- IMPORTANTE!!! É prudente digitalizar os documentos por lotes de tamanho razoável e de fazer periodicamente cópias de backups dos arquivos ".REP" e ".OUV".

11. O que fazer se...

Se você seguiu rigorosamente os conselhos apresentados neste documento e mesmo assim um incidente impedir a leitura de uma folha, eis algumas maneiras de solucionar alguns problemas mais comuns. Não esqueça que, muitas vezes, quando uma folha está apresentando problemas para ser lida, é mais rápido realizar uma digitação manual do que ficar tentando retificá-la.

11.1 O scanner não inicializa

- Verificar se o mesmo está ligado. O scanner deve sempre ser ligado antes de ligar o computador.
- Se não for este o problema, você poderá reinicializá-los juntos, ou seja, «atualizar a porta SCSI» a partir da sequência: botão Iniciar, Configurações, Painel de Controle, Sistema, Gerenciador de periféricos, Placa SCSI, Atualizar.
- Realize um teste com o programa Imaging (normalmente fornecido juntamente com o Windows e que se encontra a partir do botão Iniciar, Programas, Acessórios).
- Se a leitura não funcionar corretamente, contate o administrador responsável pelo seu computador e solicite que ele verifique se os drivers da placa SCSI ou do scanner não estão danificados.

11.2 O carregador não está ativo

Se o scanner não for um dos scanners que recomendamos, é provável que o carregador não seja 100% compatível com as normas TWAIN.

11.3 Várias páginas são carregadas ao mesmo tempo

Se as páginas estão úmidas, ou grudadas umas às outras, o carregador do scanner não terá como separá-las. Retire-as e folheie o maço das folhas, manualmente, assim como normalmente se faz quando são inseridas folhas em branco numa máquina de xerox ou mesmo numa impressora, alinhando-as e separando-as. Não esqueça que a qualidade do carregador é um forte componente do preço de um scanner.

11.4 Mensagem «Impossível detectar o quadro de página»

Verifique a imagem digitalizada, é possível que o quadro esteja com seu traço interrompido (devido à uma reprodução de pouca qualidade, por exemplo). Verifique também que nada esteja escrito sobre o traço do quadro da página.

11.5 Mensagem «Impossível de ler o código de página»

A folha pode ter sido introduzida no sentido inverso. Você poderá marcar a opção «Auto-rotação» (ver item 7.2) para forçar o sistema a testar todas as bordas da página, à procura dos códigos de página.

11.6 Mensagem «Impossível de ler a questão n»

Verificar a posição das casas esperadas e aquelas das casas detectadas (opção Áreas de Leitura). Se as casas tiverem sido apagadas, será necessário reconstituí-las. Se necessário, a posição das casas poderá ser manualmente retificada.

11.7 Mensagem «Pesquisa incompatível»

O número da pesquisa esperada é indicado na tela de leitura. Este número é o que deve aparecer na base do formulário, sob a forma de códigos de página. Se os números forem diferentes, significa que uma impressão do formulário foi realizada após a reprodução do documento. Nesses casos, é necessário retornar ao backup anterior, ou modificar manualmente (a seu próprio risco), o número identificador da pesquisa.

11.8 Exibição da tela de incoerência de imagem

- formulário pode ter sido modificado (sem ter sido impresso) depois da reprodução do documento em papel. Altere as opções das propriedades do formulário para encontrar a mesma configuração que o documento impresso.
- formulário pode ter sido realizado numa máquina (computador) diferente. Coloque a máquina em configuração padrão (fontes pequenas).
- Seja qual for a causa desta mensagem de erro, os formulários impressos poderão ser lidos por eles mesmos se a exibição na tela for diferente.

11.9 Erros de detecção das marcas

Se ao verificar as casas detectadas, for possível identificar erros entre as casas marcadas e aquelas que são vistas no formulário papel, verifique o que o scanner realmente detectou ao examinar a imagem digitalizada. Pode ser que a regulagem do contraste e da luminosidade não estejam corretos. Certas cores podem ser sido ignoradas: verifique que a caneta utilizada não seja dessa cor que foi ignorada.

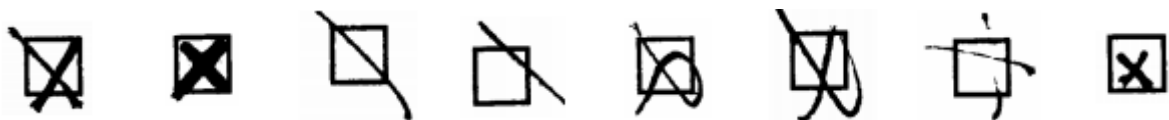
12. Algumas dúvidas mais freqüentes

*** É necessário preencher completamente as casas?**

Não, basta marcar a casa.

Se as casas forem completamente marcadas elas poderão ser reconhecidas como rasuras.

*** Como marcar uma casa? O que acontece se o traço ultrapassar a casa**?

Não há necessidade da marca ser precisa. Eis alguns exemplos (maximizados) de marcas reconhecidas:

… e algumas casas ou marcas não reconhecidas:

Por padrão, esta será considerada como uma mancha ou borrão

O traço é muito próximo do limite da casa

Por padrão, este será considerado como uma rasura

A máquina não é tão inteligente!

OBS: é possível regular o percentual de preenchimento que será considerado como uma marca efetiva.

*** Pode-se utilizar um lápis?**

Sim. Basta que o traço seja suficientemente grosso para ser identificado pelo scanner. Um traço de lápis é reconhecido sem dificuldade. Recomendamos evitar canetas com tinta de cor muito clara.

*** Pode-se rasurar?**

Sim. Se uma casa foi marcada errada, basta preenchê-la integralmente e ela não será considerada como marcada (a sensibilidade desta detecção de rasura pode ser regulada).

*** O que acontece se duas casas forem marcadas?**

O módulo detecta se a quantidade de casas marcadas é incorreta, de acordo com o tipo de questão. A página correspondente será marcada em vermelho na lista de páginas lidas. O módulo considera a marca mais provável (a mais próxima da marca ideal). A verificação e a modificação eventual podem ser efetuadas manualmente pelo usuário após a leitura de um lote de questionários.

*** Qual a confiabilidade?**

100%! Se os questionários forem preenchidos com um mínimo de cuidado, não terá nenhum erro de reconhecimento.

*** Qual a velocidade que se pode esperar?**

A velocidade depende da velocidade do scanner e da potência do computador que vai analisar as imagens digitalizadas. Por exemplo, com um computador de mesa padrão e um scanner Fujitsu M3092, tem-se uma velocidade de 300 páginas/hora (para questionários com 20 questões por página).

*** Como tratar questões abertas ou numéricas?**

Manualmente, após a leitura de um pacote de questionários preenchidos, você poderá consultar as imagens das áreas de questões abertas e, com um simples clique, você poderá digitar o texto sem necessidade de retornar ao questionário papel.

*** Pode-se «codificar» um lote de folhas?**

Freqüentemente, um lote de questionários possui características comuns (n° de serviço, data, código de formação, etc). Esta característica corresponde a uma questão aberta no questionário. Esta questão aberta não será interpretada automaticamente, mas é possível – num só clique – atribuir um mesmo valor a todos questionários do lote.

*** O questionário pode ter várias páginas?**

Sim, sem nenhum problema, o número da página aparecerá sob a forma de código binário no rodapé da página. As páginas de um mesmo questionário deverão estar em ordem no carregador. Ocorrendo uma inversão de páginas dentro de um mesmo questionário, o sistema ainda assim reconhecerá que se tratam de páginas de um mesmo questionário. Entretanto cuidado! As páginas de um questionário x não podem estar

misturadas com páginas de um questionário y. Este tipo de reconhecimento não será feito pelo sistema. As páginas ausentes (ou faltando) num questionário serão ignoradas.

*** Pode-se modificar a formatação do questionário no editor de textos?**

Normalmente, o sistema reconhecerá automaticamente somente formulários impressos por ele mesmo. Se você modificar o documento, a posição das casas será alterada e o módulo não conseguirá reconhecê-las. Normalmente, as diversas opções de paginação são suficientes para satisfazer mesmo os mais exigentes em termos de edição.

*** Pode-se pré-preencher o formulário (com o nome por exemplo)?**

Sim. Os questionários podem ser impressos com certos campos pré-preenchidos. Durante a leitura, a identificação do questionário pelo número de cliente será automática, por exemplo.

*** A instalação do scanner é complexa?**

O scanner se instala ou se conecta ao computador em uma placa (SCSI ou USB) que já deve estar instalada no computador. O scanner será então «pilotado» por um sistema que é fornecido pelo distribuidor do scanner sob a forma de um disquete ou CD e que deve estar instalado no computador. O módulo configura automaticamente seu scanner com os melhores ajustes (tamanho da imagem a digitalizar, resolução ótima, etc.), não sendo normalmente necessária nenhuma intervenção especializada.

*** Qual scanner devo escolher?**

Deve ser um scanner que contenha um carregador automático de documentos funcionando com a norma Twain 32 bits. A maioria dos fabricantes propõem scanners deste tipo adaptados ao gerenciamento automático de documentos (GEIDE). Entre em contato conosco (suporte@sphinxbrasil.com) para conhecer a lista atualizada de scanners recomendados.

*** Eu já tenho um scanner, será que posso usá-lo?**

Atenção! A maioria dos scanners normais, ditos scanners de mesa, os quais são normalmente encontrados em grande distribuição, não são adaptados ao Sphinx Ambiente Multimídia opção Scanner. Se o scanner não estiver munido de um carregador fixo de documentos, o benefício da leitura por scanner será completamente anulado pelo tempo necessário e a presença constante de alguém para colocar o questionário (cada um deles) e lançar a leitura óptica. Certos scanners permitem adicionar um carregador opcional. Em vários casos, este carregador não é 100% compatível com as normas TWAIN e funcionam somente para aplicações OCR fornecidas juntamente com o scanner. Enfim, esses scanners são perfeitos para digitalizar imagens coloridas com uma excelente resolução, mas sua velocidade operacional é muito lenta para uma aplicação destinada ao ganho de tempo e agilidade na entrada de respostas de pesquisas.